DOCUMENT RESUME

ED 166 239 TM 008 233

AUTHOR	Ward, William C.; Frederiksen, Norman
TITLE	A Study of the Predictive Validity of the Tests of Scientific Thinking. Research Bulletin RB-77-6. (May 1977). GRE Board Professional Report GREB No. 74-6P.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	GREB-77-6
PUB DATE	Oct 77
NOTE	58p.
AVAILABLE FROM	Graduate Record Examinations, Educational Testing Service, Princeton, New Jersey 08541 (free while supplies last)
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	Academic Achievement; *College Entrance Examinations; *Creative Thinking; Creativity Tests; Graduate Study; Higher Education; Item Sampling; Occupational Tests; Performance Tests; *Predictive Validity; *Psychology; Questionnaires; Scientific Methodology; Self Evaluation; *Test Validity
IDENTIFIERS	*Graduate Record Examinations; *Tests of Scientific Thinking

ABSTRACT

This study provides preliminary evidence as to the validity of measures derived from the "Tests of Scientific Thinking" (TST). The TST and the Graduate Record Examinations (GRE) were compared with regard to their relationships to interests, self-appraisals, and accomplishments of students during their first year of graduate work in psychology. Criterion variables were obtained from a questionnaire mailed to students near the end of the spring academic semester. Difficulties in data analysis were created both by the item-sampling character of the experimental test data, and by a relatively small rate of return of questionnaires. It is possible to show, however, that the correlations do exhibit appreciable nonchance variance with regard to consistency across tests and across related questionnaire variables. The GRE tests were found to be more effective than the experimental instruments in predicting quality of the department attended, but the experimental tests were more effective in two other domains: self-appraisals of knowledge of psychology and skills in psychological activities; and professional accomplishments such as research, publications, and teaching. (Author/ROF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

THE GRADUATE RECORD EXAMINATIONS

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

GRE

A STUDY OF THE PREDICTIVE VALIDITY OF THE TESTS OF SCIENTIFIC THINKING

William C. Ward

and

Norman Frederiksen

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dorothy Urban
ETS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

GRE Board Professional Report GREB No. 74-6P

Research Bulletin RB-77-6
(May 1977)

October 1977

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

TM008 233

EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY □ BERKELEY, CALIFORNIA □ EVANSTON, ILLINOIS

A STUDY OF THE PREDICTIVE VALIDITY OF THE TESTS OF SCIENTIFIC THINKING

William C. Ward

and

Norman Frederiksen

GRE Board Professional Report GREB No. 74-6P

Research Bulletin RB 77-6
(May 1977)

October 1977

Copyright © 1977 by Educational Testing Service. All rights reserved.

Abstract

The "Tests of Scientific Thinking" (TST) are free-response job-sample tests intended to measure aspects of scientific problem solving that require some degree of imaginativeness and ingenuity and that might therefore be used as dependent measures in research on "creativity." The present study was designed to provide some preliminary evidence as to the validity of measures derived from these instruments. The TST and the GRE tests were compared with regard to their relationships to interests, self-appraisals, and accomplishments of students during their first year of graduate work in psychology. These criterion variables were obtained from a questionnaire mailed to students near the end of the Spring, 1975, academic semester. Subjects were students who had taken the tests in October of 1973 for a study of the psychometric properties of the experimental tests.

Difficulties in data analysis were created both by the item-sampling character of the experimental test data, and by a relatively small rate of return of questionnaires from students who were actually attending graduate school in psychology. The estimated correlations for individual variables, therefore, reflect a substantial amount of error and can be regarded only as suggestive. It was possible to show, however, that the correlations do exhibit appreciable nonchance variance with regard to consistency across tests and across related questionnaire variables.

The means, standard deviations, and reliabilities of the TST scores were in general about the same as those obtained in the psychometric study, as were correlations with GRE scores. The GRE tests were found to be more effective than the experimental instruments in predicting quality of the department

A Study of the Predictive Validity of the Tests of Scientific Thinking*

The "Tests of Scientific Thinking" (TST) are free-response job-sample tests which simulate tasks that might be encountered by a behavioral scientist. These instruments are intended to measure aspects of scientific problem solving that require some degree of imaginativeness and ingenuity and that might therefore be used as dependent measures in research on "creativity." In a previous report (Frederiksen & Ward, 1975) the tests and associated scoring methods were described, and evidence as to their psychometric properties was presented. The current report contributes additional evidence which is needed before the tests are employed as criteria in studies of creativity.

Three kinds of validity are generally distinguished: content validity, construct validity, and predictive validity. Content validity is demonstrated by showing that a test elicits behaviors that are representative of the domain of criterion behaviors. In the case of tests of scientific thinking, we lack the detailed knowledge of the criterion domain to make such a demonstration, except in an impressionistic sense (which may be little different from "face" validity). Construct validity is demonstrated by showing that correlations of the test scores with a wide variety of other measures are consistent with theoretical, or at least intuitive, expectations. A construct validity study with these tests is in progress and will be reported separately. Predictive validity, finally, requires (a) that one or more acceptable criterion variables be obtainable and (b) that the test being validated be administered well before the criterion behavior is exhibited. The present study is a predictive study

*This research was supported by the Graduate Record Examinations Board.

at least in the sense of the time relationships in data collection. The criterion variables were chosen because they appear to be relevant to our evaluation of the Tests of Scientific Thinking, but they can in no sense be considered validated measures of imaginative problem solving. In this sense, the study is more like construct validation, since the "criterion" variables are, in varying degrees, intuitively relevant to the question of what the tests measure.

Summary of Previous Study

The four tests of scientific thinking (Frederiksen & Ward, 1975) are Formulating Hypotheses (FH), Evaluating Proposals (EP), Solving Methodological Problems (SMP), and Measuring Constructs (MC). Candidates are instructed to propose one or more solutions to each problem posed by the test items; they are asked to write not only the answer they consider best but also others that they think deserve consideration. A scoring method is used in which the scorer, rather than making a subjective evaluation, assigns each answer to one of the categories in a classification of answers to each problem. These categories are given scale values based on rankings by a panel of judges. The method permits six scores to be generated for each item: three are concerned with the quality of solutions offered--quality of Best response (based on candidate's choice of his best answer), Mean quality of responses, and Highest quality response (based on the scale values); and three depend on counts of the number of ideas given--Number of responses, number of Unusual responses, and number of responses that are both Unusual and High in quality.

The tests were administered as part of a regular administration of the GRE Advanced Psychology Test, using an item-sampling procedure. The analysis was intended primarily to evaluate the tests from the standpoint of their psychometric properties. Scores from six-item tests were sufficiently reliable to justify their use in research, except that the Unusual-High Quality score was unsatisfactory in the case of SMP and MC. The tests were judged to be of suitable difficulty, although SMP was found to be a bit too difficult for the GRE candidate population. The standard error of measurement for a more select subsample was somewhat smaller than was that for the entire group, indicating that the tests could be used with more advanced students.

A factor analysis of the intercorrelations of all the scores generated by the TST revealed three factors: A number-of-responses factor including number scores from all four tests, and two quality factors, one based on quality scores from FH and MC and the other on quality scores from EP and SMP.

Correlations with GRE scores were low, even when corrected for unreliability; number scores were even less closely related to GRE scores than were the quality scores. It was concluded that a substantial proportion of the true variance in the experimental test scores is not accounted for by present GRE tests, and that the new tests warrant further study, particularly of their predictive and construct validity.

Method

The availability of TST scores for a sizable cohort of candidates for admission to graduate school suggested the desirability of a follow-up study as one approach to the investigation of validity. The study was based primarily on 1,600 candidates, each of whom took a three-item version of one of the

tests in accordance with the item-sampling plan. Each candidate was asked to respond to a questionnaire whose items elicited information about type of school and program attended and about interests, plans, self appraisals, and accomplishments during the first year of graduate study.

Thus, the purpose of this investigation was to discover relationships of scores on the experimental tests to items of information obtained from the questionnaire that are relevant to test validity, to compare these relationships with those involving the GRE tests, and to see to what extent the new tests might make independent contributions to the prediction of accomplishments as indicated by questionnaire responses.

It was recognized that self-reports based on questionnaire items after only one year of graduate work are inadequate as criteria of success, and that simple dichotomizations cannot fully capture the valid variance on which answers to questionnaire items are based. Furthermore, the assumptions involved in the item-sampling procedure are not likely to be fully satisfied for the follow-up sample. The most important of these assumptions is that the subgroup taking each of the many subtests required for item sampling is a random sample, and thus representative of the total group. This assumption is more likely to be violated as selection, including self-selection, operates to reduce the size of the subgroups. Despite these difficulties, the number and diversity of individuals available for study make this sample a unique resource with which to search for relationships. At the least, it was hoped, suggestive evidence should be obtained which might later be confirmed through studies involving administration of full-length tests to more homogeneous groups of students.

The Questionnaire

A copy of the "GRE Questionnaire" is included as Appendix A. The first six items, intended for all recipients of the questionnaire, made it possible to compare groups that differed with regard to graduate school attendance, type of graduate program, and reasons for nonattendance.

The remaining items were to be answered only by those who attended a graduate program in psychology during the 1974-75 academic year. These items were concerned primarily with intended areas of specialization, activity preferences, self-appraisals, and accomplishments during the year. Candidates were also invited to report the name of the university they were attending, which made it possible to generate variables describing the quality of the program. These variables are based on published information concerning accreditation (APA, 1974) and departmental evaluations (Roose & Andersen, 1970). The student reports of their accomplishments (Item 17) are perhaps the most important from the standpoint of predictive validity.

Forty-two variables were generated from the questionnaire responses. Table 1 provides the names of the variables, the range of values possible for

---
Insert Table 1 about here
---

each, the questionnaire item (or items) on which each is based, and the method of translating responses into a numerical code. All are coded so that high numerical values represent the positive or "high" end of the scale.

In this and subsequent tables, the questionnaire variables have been grouped into six general areas: indices of student and department quality; program emphasis; areas of professional interest; preferred professional

activity; self-appraisals; and professional activities and accomplishments during the academic year.

Description of the Sample

The number of complete-data cases obtained in the previous investigation was 3,586. The sample reduction that occurred at various stages is shown in Table 2.

Insert Table 2 about here

The returned questionnaires were sorted into three categories. Category 0 contained questionnaires completed only through Item 6; students in this category either did not attend graduate school or were in fields other than psychology proper (including, for example, guidance and counseling). Category 1 subjects did attend graduate school in psychology; these were the subjects used in the main analysis. Category 2 subjects attended graduate school, not in psychology but in fields related to psychology, but nonetheless completed the entire questionnaire.

As Table 2 shows, questionnaires were mailed to approximately 90% of the original complete-data sample, and presumably received by 75% of that group. Fifty percent of those who received questionnaires returned them completed. However, of those who returned questionnaires, only 49% met the criteria for inclusion in the main group; 44% did not attend graduate school, or attended in an area unrelated to psychology, while 7% were in fields only related to psychology. The number of Category 1 students was 654. Of these, 403 were from the intercorrelation sample in the previous study (those who took one item

from each of two different tests) and 251 were in the reliability sample (those given three items from a single test). This latter sample is the most useful in the present study, since for these individuals it is possible to estimate relations based on three-item rather than one-item tests. Thus the various sources of reduction combined to leave available data that are sufficiently abundant to justify analysis but probably not representative of the original group.

Table 3 shows the means and standard deviations of GRE scores for the

---

Insert Table 3 about here

---

three groups of students who returned questionnaires as well as for several subgroups that did not provide data. Among those who did respond, Category 1 students--those who attended graduate school in psychology--were clearly the most capable in terms of GRE scores, with means only a little below 600. Category 0 students, who either did not attend or attended a nonpsychological program, were substantially lower, and those in Category 2 (programs related to psychology) were still lower. The differences are substantial for all the GRE scores, but are smaller for GRE-V than for GRE-Q and the Advanced Psychology Test, probably indicating a lesser emphasis on the use of the Verbal test in graduate school selection.

Among students on whom follow-up data were not obtained, the means for the "Not Deliverable" group are only a little lower than the average for all those who returned questionnaires. The remaining two groups--those who presumably received questionnaires but chose not to return them, and those ("Not Sent") who did not grant permission to be contacted in the follow-up--were

substantially lower, indicating some self-selection on ability among those who cooperated in the study.

The last column of Table 3 contains mean scores for the original GRE sample. The main follow-up sample had substantially higher means (as might be expected, since GRE scores are used in selection of graduate students), but was not notably restricted in the range of scores.

Table 4 shows the means and standard deviations of TST scores based on

---
Insert Table 4 about here
---

data from the original GRE study and from the follow-up. Unlike results involving the GRE tests, where the follow-up sample has substantially higher average scores, the experimental test means for the two groups are quite similar. Only the GRE scores were used in selection. But there is a tendency for SMP means to be higher for the follow-up group, which is consistent with the hypothesis that the difficulty level of this test is more appropriate for the better or more advanced students.

A description of the sample in terms of the questionnaire variables is presented in Table 5. Table 1 should be consulted in studying this table

---
Insert Table 5 about here
---

because it provides information on how the variables were constructed from the questionnaire items. Some highlights: Students in the main sample studied (Category 1) tend to aim for a Ph.D.; to earn high grades; to prefer applied research, clinical practice, or teaching; to prefer clinical practice when asked to choose one area; to rate themselves high on knowledge of psychology

and statistics and low on teaching skills; to have a fellowship or assistantship; and to have had more experience in research activities than in writing and editing during their first year of graduate study.

Table 5 also includes means and standard deviations on questionnaire items for Category 2 students, together with t tests showing the significance of differences between the two groups. Many differences are highly significant. The students who are in a field related to psychology reported lower undergraduate GPA's on the average, but similar graduate school GPA's. The Category 2 students less often planned to earn a Ph.D., were more likely to be in an applied area, named clinical practice and counseling as preferred activities more often and research less often, rated themselves lower in all the skills areas except clinical, and claimed fewer accomplishments during the first year of graduate study.

Obviously a great deal of selection has taken place in going from the sample employed in the original investigation to that available for the present study. The reliability sample, which provides the best data for present purposes, constitutes only 19% of the original GRE reliability sample. One result is considerably higher GRE scores and another, apparently, is students who are more serious about a career in psychology.

Reliabilities of Scores for Six-Item Experimental Tests

Reliabilities of the scores from each of the four experimental tests were calculated using data from subjects in the follow-up study who had been members of the original reliability sample. The coefficients are presented in Table 6,

---

Insert Table 6 about here

---

along with the comparable estimates reported for the same scores in the previous study. These are lower-bound estimates of the reliability of a six-item test based on "Lambda 2" (Guttman, 1945); their computation is described in the statistical appendix to the original report.

With three exceptions, the reliabilities are higher for the follow-up study, in some cases substantially so. The mean difference in coefficients between the two studies is .11, while the median difference is .08. This result might be interpreted to indicate that the experimental tests are more appropriate for the more highly selected students contributing data to the follow-up study. However, it is also possible that this result arises from the greater variability in the follow-up data of the terms used in calculating the coefficients, especially since the computation of Lambda 2 involves the use of squares of interitem covariance terms. Both the decreased representativeness and the smaller N's of the subsamples in the follow-up data might lead to their having greater variability and thus to inappropriately high estimates when their squares are entered into the formula.

In Table 7 are given coefficient alpha reliability estimates for the two

---

Insert Table 7 about here

---

studies. The computation of alpha involves the use of interitem covariances rather than their squares, and thus should be less heavily influenced by instability in the component terms. On the average, these estimates for the two studies are quite similar. Their means differ by .01, and their medians by .03; 13 are higher and 10 are lower in the follow-up data.

In view of the rather erratic variation between the coefficients obtained in the two studies, one should probably withhold judgment about the reliability of any single score; but it appears that the experimental tests in general are as reliable for the follow-up sample as for the original group.

Correlations of GRE Scores with TST Scores

Correlations between experimental test scores and the GRE scores are given in Table 8 for the follow-up study data and, for comparison, for the data from

---

Insert Table 8 about here

---

the original GRE study. In general, the estimates are similar for the two studies. Across all experimental test scores and all GRE scores, coefficients are slightly lower for the follow-up study, with 44 of the 115 coefficients higher and 67 lower. The mean difference in coefficients is .03 and the median difference is .02. There is some tendency for correlations involving the number scores to be higher and correlations involving quality scores to be lower for the follow-up sample. But SMP is a striking exception; for SMP the quality score correlations are higher. These are the scores that in the original study had the highest relation to GRE tests and therefore might have been most susceptible to restriction of range because of indirect selection. The fact that these correlations turned out to be higher rather than lower than those originally obtained perhaps suggests that SMP functions relatively better than the other tests when a more select sample of students is involved.

Correlations of GRE Scores with Questionnaire Variables

Table 9 shows the correlations between GRE scores and the 42 variables derived from the questionnaire. Sample size and item sampling are not matters

---

Insert Table 9 about here

---

of concern here, since all individuals in Category 1 provide data for the computation of these coefficients. The N's approximate 650 for correlations with the Advanced Psychology Test and its two subscores, Experimental and Social-Personality, and 525 for correlations with the GRE Verbal and Quantitative Aptitude scores.

Fifty-five percent of the coefficients in the table are significant at the 1% level. The Advanced Psychology Test and its Experimental subscore are in general the two best predictors, but the five GRE scores tend to be quite similar with regard to their correlations with questionnaire variables.

An examination of correlations by area shows that the GRE scores are by far most consistently related to the loose grouping of questions which we have identified as indices of student or department quality. These include questions having to do with the student's degree aspirations and academic achievement, with ratings of department quality, with indirect indices such as whether or not the student received fellowship support during the year, and even with reports of satisfaction with the program. Forty-two of the 45 coefficients in this area are significant at the .01 level. Of particular interest are the relations of GRE scores to graduate grade-point average, since this is the single most frequently used criterion for the validation of selection instruments.

The correlations range from .18 to .26, with the highest being that involving the Advanced Psychology Test.

In several additional areas the GRE scores show a substantial number of significant correlations, though without such overwhelming consistency. In terms of students' desired professional activities, they are significantly related to the number of activities checked, to interest in both applied and basic research, and to a lack of interest in guidance and counseling. In self-appraisals of skills and knowledge relevant to psychology, they predict particularly self-reported ability to evaluate and interpret research. Also, higher GRE scores are associated with attending a program whose objectives are academic or research-oriented rather than to prepare practitioners.

Finally, there are two areas in which only a scattering of significant coefficients appear. In the domain of the preferred professional activity, only a lack of interest in guidance and counseling shows consistent relations to the GRE scores. In that of professional activities and accomplishments during the academic year, none of the questionnaire variables is consistently related to the GRE scores.

Along with the large number of statistically significant correlations, however, it is important to note the generally low level of prediction obtained from the GRE scores. With the large N's available, a "significant" correlation may be no larger than .10.

To facilitate comparison of the GRE and TST scores as predictors, we have underlined all those correlations in the table which are .25 or higher. This figure was chosen since, with the N's typical for the experimental test correlations to be reported below, a coefficient of .25 is required for significance at

the 5% level. Only 17 correlations between GRE scores and questionnaire variables equal or exceed this cut-off level. Those are concentrated, with two exceptions, in the portion of the table representing indices of student and department quality. Further, the bulk of these (10 of 15) have to do with the two objective indices of department quality--whether or not the department has APA accreditation and the Roose-Andersen rating--and might indicate only the effects of a selection process in which students scoring higher on the GRE tests are more likely to be accepted by better psychology departments. All in all, then, if the question is one of substantial rather than merely statistically significant prediction, the GRE scores show very little in relation to student interests and activities during the first year of graduate school in psychology.

Relations of TST Scores with Questionnaire Variables

It was originally intended that correlations between TST scores and variables derived from the questionnaire would be computed using the estimation procedures for item-sampled data described in the earlier report on these tests (Frederiksen & Ward, 1975). These procedures require the use of item variance and covariance terms analogous to those employed in reliability estimation. However, as was seen above, reliability coefficients for the experimental test scores were somewhat unstable, presumably because of small subsample sizes coupled with the likelihood that the assumption of random assignment to item-sampling subgroups does not hold. It is therefore probable that the estimated correlations would also be unstable. Moreover, no test of significance would be available, since the conventional formula for the standard error is not appropriate for these coefficients.

Because of these problems, an alternative method for computing correlations was chosen in which it was not necessary to use the item variance and covariance terms. The reliability sample provides the appropriate data. Each subject in that sample took a subtest composed of three items from one test. Since there are 20 possible three-item combinations of six items, very few subjects took exactly the same test; but if it is assumed that the items are interchangeable (after adjusting for differences in item means), scores can be obtained and correlations computed based on whatever three-item test was taken. The resulting coefficients may be subject to less fluctuation than those derived from the estimation procedure, and in any case they can be tested for significance by conventional methods. The correlational results to be reported below are based on such three-item tests.[1]

Each item score was adjusted by subtracting the grand mean obtained from the complete reliability sample in the original study. Since sample sizes and item-sampling methods in that study give reasonable assurance of equivalence of ability for subgroups, this procedure should provide an appropriate adjustment for differences in item difficulties. After this adjustment, six test scores were obtained for each subject by summing over the three items from the test he had taken. These scores were then correlated with the 42 questionnaire variables.

The results are presented in Tables 10 through 15. In each table are

---

Insert Tables 10 through 15 about here

---

given correlations for one of the six scores (e.g., quality of "Best" response) from all four of the tests. Coefficients significant at the .05 level are underlined.

A count of the number of coefficients reaching the .05 level of significance is revealing. Over the three tables of correlations with quality scores, a total of 21 coefficients out of the 462 displayed attain this level of significance. Since the expected number of correlations "significant" by chance is 23, it is impossible to conclude that any real relations are represented in these tables. For the remaining three tables--scores for Number, number of Unusual, and number of Unusual-High quality responses--the situation is slightly better; here 37 of 504 coefficients, or 7.3%, reach significance at the 5% level. There is evidently some nonchance variation in this part of the data; but it is heavily embedded in noise, and a coefficient-by-coefficient discussion of these results could not be justified.

Note, however, that the counts just presented are for correlations based on three-item tests. In all our work to date, we have assumed that in using these tests we would employ six items in order to have more adequate reliability while remaining within a reasonable time requirement. With the greater reliability of a longer test, a larger number of significant coefficients would be expected. The procedure for estimating correlations for a six-item test, described in Frederiksen and Ward (1975), was applied to the data, and the number of resulting coefficients of .25 or greater was examined. For the quality scores, 29, or 6.3%, were of this magnitude; for the count scores, 76, or 15.1%, were this large. These results again suggest a lack of real relationships between these indices of graduate school performance and indices of quality on the experimental tests, while those scores which depend on counts of numbers of ideas do show more relations than can be attributed to chance.

Tests of the significance of the combined results for the four tests were made for average correlations based on r to z transformations (McNemar, 1962). The four correlations in any row of the preceding tables may be viewed as four replications of a relationship; they are independent replications in the sense that the four coefficients are based on nonoverlapping samples of subjects. The three scores based on counts (Number, Unusual, and Unusual-High) are considered together because the factor analysis performed in the earlier study showed that they loaded on one factor, in contrast to quality scores which loaded on two factors.

Table 16 provides a summary of significant coefficients obtained by this procedure. Number, number of Unusual, and number of Unusual-High quality

---

Insert Table 16 about here

---

responses show, respectively, 7, 9, and 10 relationships significant at at least the .05 level, where 2 for each score would be expected by chance; so it is possible to examine those that appear in the table with assurance that most of them are real.

There is evident consistency across the three scores in the location of significant results within sections of the table. Of the 26 average correlations significant at at least the 5% level, 12 are derived from questions we have grouped under "Professional Activities." Students scoring high on the experimental tests tend to have engaged in more of these activities, and specifically to have attended professional meetings and subscribed to journals, published, engaged in collaborative research, provided advice on experimental design and statistics to other students, and worked with laboratory equipment.

Eight significant relationships are found within questions grouped under "Student/Department Quality." Students scoring high tend to have plans for the Ph.D. rather than a master's degree, to attend a department with a high Roose-Andersen quality index, and to have obtained their support through a psychology-related activity during the first year. Four more coefficients are found within the section of the table designated "Self-Appraisal." Students scoring high on number of Unusual responses generally rate themselves high; along with those high on Number of responses, they claim knowledge of experimental design; and those high on number of Unusual-High quality ideas see themselves as having a lower level of clinical skills. Finally, there are significant relationships between number of Unusual-High quality ideas and two indices indicating enrollment in a program which emphasizes research rather than practice.

A similar analysis was carried out on correlations of questionnaire variables with the three quality scores. In view of the chance number of significant relationships that had been seen for individual correlation coefficients, it was not expected that positive results would have been found. And, in fact, only four relationships reaching the .05 level were found, where the chance expectation would have been six.

These results indicate some ability for certain of the TST scores to predict possibly important aspects of first-year graduate performance, in spite of their low reliability. Additionally, they suggest a discrimination between the experimental tests and the conventional predictors as to what aspects of this performance are predicted. Recall that, in examining correlations between the Advanced Psychology test and the questionnaire variables,

substantial relations were found mainly with variables representing individual or departmental quality indices, while there was little relation to variables representing professional activities.

Discussion and Conclusions

In some respects the present data were disappointing. We had hoped, perhaps naively, for a larger return; 3,200 individuals had indicated at the time of the original GRE testing that they were willing to be contacted in a follow-up. We had expected that, among those who did respond, a larger proportion would have met the criteria for inclusion in the main analyses to be conducted; of those who did cooperate with the study, only 49% were graduate students in psychology during the past year. The difficulties in analysis raised by the loss of cases, particularly in view of the problems created by the item-sampling origin of the data, are abundantly clear in the presentation of results. Our best validity information is based on the average validity of four three-item tests, and the three items not the same for all subjects. The reliability of such tests would be substantially lower than those reported for six-item tests.

Given these severe limitations in the data, the relationships which do emerge are perhaps more encouraging than could have been expected. We find that the experimental measures have some ability to predict indices of student and departmental quality, the area in which the GRE tests are most effective as predictors. In addition, there is evidence that the experimental tests are related to two clusters of variables--self-appraisals of professional skills and reports of professional activities and accomplishments--which have some face validity as potential additional criteria for predictive efforts, but which are

not strongly related to the GRE scores. Moreover, the best scores from among the six experimental scores, in terms of apparent predictive power, are those for the number of Unusual responses and the number of Unusual-High quality responses, which are in general the least reliable of the experimental test scores. Since these are the scores from the TST which have the lowest relations to GRE scores, the results suggest that whatever predictive validity the experimental tests have is incremental rather than arising from an overlap with the psychology achievement test.

It should also be noted that the questionnaire variables which served as criteria against which to validate the experimental test scores are themselves, in many cases, highly tentative and exploratory. Some of these variables, such as graduate grade-point average and number of publications, are clearly relevant indices of the quality of the student's performance in graduate work. Others, however, are indirect in their import--they are more like indices of styles or preferences than of level of performance, as in the areas of professional interest. An absence of significant relations with many of these latter variables cannot be labeled a failure of either the conventional or experimental tests. In the various analyses in which we have counted the number of apparently significant coefficients and compared these with the number expected over all 42 questionnaire variables, therefore, we have to some extent erred on the conservative side in the claims made for the test scores.

Going one step further, the data obtained from this questionnaire cannot be considered as exhausting the set of relevant criteria against which the test scores might be validated. Other aspects of first year graduate performance might be considered, and some of the aspects which were assessed could be measured more sensitively than by self report. Further, the first year of graduate

work is itself likely to be quite inadequate as an index of future scientific performance.

For a variety of reasons, then, we have much less evidence than is needed to judge what the long-run potential usefulness of these instruments will be. The present study suggests some kinds of performance for which the tests are likely to be especially relevant, as well as some differences among the tests and their various scores as to which are likely to be most effective. These suggestions will have to be replicated in more tightly controlled investigations which avoid some of the difficulties inherent in the design under which this exploratory work was performed. We remain hopeful that the tests will eventually prove valuable as instruments for use in basic experimental studies, as prototypes for possible new item types for inclusion in the GRE tests, and potentially for use in the training of students and in the evaluation of instructional programs.

References

American Psychological Association. APA-approved doctoral programs in clinical, counseling, and school psychology: 1974. American Psychologist, 1974, 29, 844-845.

Frederiksen, N., & Ward, W. C. Development of measures for the study of creativity. Research Bulletin 75-18. Princeton, N. J.: Educational Testing Service, 1975.

Guttman, L. A basis for analyzing test-retest reliability. Psychometrika, 1945, 10, 255-282.

McNemar, Q. Psychological statistics. (3rd ed.). New York: Wiley, 1962.

Roose, K. D., & Andersen, C. J. A rating of graduate programs. Washington, D. C.: American Council on Education, 1970.

Footnotes

[1]This procedure sacrifices the opportunity to use intercorrelation sample data to provide a partly independent replication of results obtained with reliability sample data. However, a preliminary analysis, in which the signs of correlations between items and criterion variables were examined for consistency, showed a difference in results for the two samples--many more apparently significant relationships to the criteria than would be expected by chance were found in the reliability sample data, but only a chance number appeared in the intercorrelation sample data. The reason for this difference is not clear. One might speculate that, since the students in the reliability sample worked on three items from a single test, while the other sample had to read new instructions after completing one item, the first group may have had a better opportunity to develop an appropriate set or strategy for attacking those problems, which in turn might have made their scores more valid.

Table 1

Questionnaire Variables

	Questionnaire Variable	Range	Question Number	Comments
	Student/Dept. Quality			
2.	Plans PhD rather than MA	0-1	4	
3.	Undergraduate GPA	2.17-4.5	5	X=#1 to 8, recoded Y=4.883-.333X
4.	Graduate GPA	2.17-4.5	6	X=#1 to 8, recoded Y=4.833-.333X
5.	Attends accredited dept.[a]	0-1	15	
29.	Rates dept. high in quality	1-5	13	X=#1 to 5, recoded Y=6-X
30.	Satisfied with department	1-3	14	X=#1 to 3, recoded Y=4-X
31.	Department quality index[b]	1-5	15	
32.	Support through psychology	0-1	16	1=#1 to 4, 0=#5
33.	Support through fellowship[c]	0-1	16	1=#1, 0=#2 to 4
	Program Emphasis			
6.	Plans academic program	0-1	7	1=#2,4, 6, 7, 9, or 11; 0=#1,3, 5,8, or 10
27.	Program objective: practice	0-1	11	1=#3, 0=#1
28.	Deemphasizes teaching	0-1	12	1=#2, 0=#1 or 3
	Interest Areas			
7.	Number of areas	0-8	8	Number of activities checked
8.	Administration	0-1	8a	
9.	Applied research	0-1	8b	
10.	Basic research	0-1	8c	
11.	Clinical practice	0-1	8d	
12.	Guidance and counseling	0-1	8e	
13.	Teaching	0-1	8f	
	Preferred Area			
14.	Administration	0-1	9-1	
15.	Applied research	0-1	9-2	
16.	Basic research	0-1	9-3	
17.	Clinical practice	0-1	9-4	
18.	Guidance and counseling	0-1	9-5	
19.	Teaching	0-1	9-6	
	Self-Appraisal			
20.	Mean rating	1-5	10	Variables 21 to 26
21.	Knowledge of psychology	1-5	10abc	X=#1 to 5, recoded Y=5-X
22.	Knowledge of statistics	1-5	10de	X=#1 to 5, recoded Y=5-X
23.	Experimental design	1-5	10fh	X=#1 to 5, recoded Y=5-X
24.	Research interpretation	1-5	10ik	X=#1 to 5, recoded Y=5-X
25.	Clinical ability	1-5	10g	X=#1 to 5, recoded Y=5-X
26.	Teaching ability	1-5	10j	X=#1 to 5, recoded Y=5-X
	Professional Activities			
34.	Number of activities	0-15	17	Number of activities checked
35.	Meetings, subscriptions	0-1	17ab	
36.	Publications	0-1	17cde	

Table 1 (continued)

	Questionnaire Variable	Range	Question Number	Comments
37.	Planned, did indep. research	0-1	17fg	
38.	Did collaborative research	0-1	17h	
39.	Taught undergraduates	0-1	17ij	
40.	Advised on statistics	0-1	17k	
41.	Helped prepare book	0-1	17l	
42.	Worked with equipment	0-1	17mno	
	Other			
1.	Has MA rather than BA	0-1	3	

[a]APA accreditation is in clinical, counseling, and/or school psychology.

[b]Ns are smaller because about 15% of the subjects did not indicate the school they are attending. Roose-Andersen rating: 5 = highest, 1 = not rated.

[c]Ns are approximately 40% smaller because the "other" category was eliminated.

Table 2

Numbers and Percentages for Complete-Data and Questionnaire Samples

	N	% of Complete Data Sample	% of Mailed Sample	% of Delivered Sample	% of Returned Sample	% of Psychology-Graduate Sample
Complete Data Sample	3586	100.0				
No Permission to Follow-up	330	9.2				
Could Not Be Matched to GRE Data Tapes for Address	21	.6				
Total Mailed	3235	90.2	100.0			
Returned as Undeliverable	546	15.2	16.9			
Presumably Delivered	2689	75.0	83.1	100.0		
Not Returned[a]	1366	38.1	42.2	50.8		
Returned	1323	36.9	40.9	49.2	100.0	
Answered Items 1-6 Only (Cat. 0)	579	16.1	17.9	21.5	43.8	
Answered All Items; Graduate Student in Psychology-Related Field (Cat.2)	90	2.5	2.8	3.3	6.8	
Answered All Items, Graduate Student in Psychology (Cat. 1)	654	18.2	20.2	24.3	49.4	100.0
Intercorrelation Sample	403	11.2	12.5	15.0	30.5	61.6
Reliability Sample	251	7.0	7.8	9.3	19.0	38.4

[a]Including approximately 15 questionnaires received after analyses had begun.

Table 3

Mean GRE Scores for Groups of Returned and Non-returned Questionnaires

GRE Scores		Returned[a]			Not Returned		Not Sent	Original GRE Sample
		Cat. 0	Cat. 1	Cat. 2	Not Deliverable	No Response		
GRE-V	M	562	585	538	561	544	552	558
	SD	96	95	100	98	99	105	100
	N	494	527	72	465	1,154	282	2,998
GRE-Q	M	554	585	504	557	533	540	549
	SD	114	108	125	114	120	117	118
	N	494	527	72	465	1,154	282	2,998
GRE-Adv.	M	546	591	531	558	539	539	552
	SD	87	88	98	92	92	92	93
	N	578	653	90	546	1,357	332	3,560
Exp.	M	54.4	58.5	52.0	55.4	53.6	53.5	54.9
	SD	8.9	9.1	9.9	9.4	9.4	9.8	9.5
	N	578	653	90	546	1,357	332	3,560
Soc.-Pers.	M	54.2	58.2	53.8	55.2	53.9	54.5	55.0
	SD	8.5	8.7	9.2	9.0	9.0	8.9	9.0
	N	578	653	90	546	1,357	332	3,560

[a]Cat. 0 = nonattendance or not in psychology.
Cat. 1 = graduate student in psychology.
Cat. 2 = graduate student in psychology-related field.

Table 4

Means and Standard Deviations of TST Scores[a]

Score		FH		EP		SMP		MC	
		Follow-up Sample	Original Sample	Follow-up Sample	Original Sample	Follow-up Sample	Original Sample	Follow-up Sample	Original Sample
Best	M	19.4	19.7	--[b]	--	16.0	14.9	14.4	14.6
	SD	3.9	3.4	--	--	2.0	2.2	4.4	3.6
Mean	M	17.8	17.9	18.0	17.8	14.6	14.0	13.0	13.7
	SD	3.0	3.0	1.7	1.9	1.9	2.0	3.1	3.0
Highest	M	22.2	22.1	24.2	23.9	17.8	17.2	17.4	18.0
	SD	3.3	2.9	1.6	1.8	2.1	1.8	3.6	3.2
Number	M	2.5	2.5	4.2	3.9	2.5	2.4	3.0	2.8
	SD	.6	.6	.9	.9	.6	.6	.9	.8
Unusual	M	.6	.6	1.2	1.1	.6	.5	.9	.8
	SD	.3	.4	.5	.5	.3	.3	.4	.4
Unusual-High	M	.3	.2	.2	.2	.2	.1	.2	.2
	SD	.2	.2	.2	.2	.1	.1	.2	.2
N		57	359	65	339	66	309	63	340

[a]Based only on "Reliability Sample"--those individuals completing three items from one of the experimental tests.

[b]No "Best" score was obtained for EP.

Table 5

Questionnaire Means and S.D.'s for Graduate Student Groups

	Questionnaire Variable	Category 1			Category 2			
		M	SD	N	M	SD	N	t
	Student/Dept. Quality							
2.	Plans PhD rather than MA	.89	.31	598	.75	.44	68	2.55*
3.	Undergraduate GPA	3.92	.47	637	3.80	.46	89	2.30*
4.	Graduate GPA	4.20	.33	624	4.21	.27	87	.31
5.	Attends accredited dept.	.51	.50	552	.57	.50	79	1.00
29.	Rates dept. high in quality	3.06	1.21	630	2.91	1.18	85	1.10
30.	Satisfied with department	2.14	.66	633	2.17	.69	89	.39
31.	Department quality index	1.99	1.39	541	2.34	1.60	79	1.85
32.	Support through psychology	.62	.48	638	.44	.50	90	3.21**
33.	Support through fellowship	.35	.48	398	.28	.45	40	.93
	Program Emphasis							
6.	Plans academic program	.37	.48	591	.05	.21	85	10.62***
27.	Program objective: practice	.50	.50	633	.91	.24	85	12.40***
28.	Deemphasizes teaching	.56	.50	633	.39	.49	88	3.04**
	Interest Areas							
7.	Number of areas	2.41	1.04	640	2.17	1.20	90	1.80
8.	Administration	.14	.34	640	.18	.38	90	.95
9.	Applied research	.50	.50	640	.23	.43	90	5.46***
10.	Basic research	.35	.48	640	.10	.30	90	6.78***
11.	Clinical practice	.55	.50	640	.48	.50	90	1.24
12.	Guidance and counseling	.24	.43	640	.68	.47	90	8.40***
13.	Teaching	.50	.50	640	.32	.47	90	3.37***
	Preferred Area							
14.	Administration	.03	.18	640	.02	.15	90	.58
15.	Applied research	.18	.38	640	.03	.18	90	6.20***
16.	Basic research	.15	.36	640	.01	.11	90	7.63***
17.	Clinical practice	.42	.49	640	.38	.49	90	.73
18.	Guidance and counseling	.08	.28	640	.44	.50	90	6.68***
19.	Teaching	.13	.33	640	.11	.32	90	.55
	Self-Appraisal							
20.	Mean rating	2.46	.50	640	2.32	.61	90	2.08*
21.	Knowledge of psychology	2.58	.59	640	2.50	.60	90	1.19
22.	Knowledge of statistics	2.40	.83	640	2.12	.95	90	2.66**
23.	Experimental design	2.43	.84	639	2.15	1.00	90	2.53*
24.	Research interpretation	2.55	.74	638	2.19	.88	90	3.70***
25.	Clinical ability	2.31	1.41	630	3.07	1.00	90	6.36***
26.	Teaching ability	2.20	1.24	636	2.06	1.25	90	1.00
	Professional Activities							
34.	Number of activities	4.05	2.64	640	2.42	2.03	90	6.85***
35.	Meetings, subscriptions	.41	.37	640	.41	.37	90	0.00
36.	Publications	.11	.22	640	.05	.16	90	3.16**

Table 5 (continued)

Questionnaire Variable	Category 1			Category 2			
	M	SD	N	M	SD	N	t
37. Planned, did indep. research	.46	.40	640	.33	.39	90	2.95**
38. Did collaborative research	.44	.50	640	.14	.35	90	7.17***
39. Taught undergraduates	.26	.38	640	.14	.29	90	3.52***
40. Advised on statistics	.24	.43	640	.17	.37	90	1.65
41. Helped prepare book	.06	.23	640	.03	.18	90	1.43
42. Worked with equipment	.24	.30	640	.06	.15	90	9.11***
Other							
1. Has MA rather than BA	.18	.39	636	.33	.47	89	2.88**

*$p < .05$ (all two-tailed tests)

**$p < .01$

***$p < .001$

Table 6

"Lambda 2" Reliabilities of Six-Item TST Scores[a]

Test	Study	Score					
		Best	Mean	Highest	Number	Unusual	Unusual-High
FH	Follow-up	.76	.70	.85	.68	.27	.48
	Original	.50	.62	.65	.67	.53	.46
EP	Follow-up	--[b]	.69	.53	.68	.74	.46
	Original	--	.61	.51	.73	.55	.44
SMP	Follow-up	.48	.58	.70	.68	.62	.37
	Original	.30	.46	.35	.61	.42	.05
MC	Follow-up	.81	.85	.78	.86	.32	.63
	Original	.68	.77	.71	.77	.36	.17

[a] *N*'s for the follow-up study are 57, 65, 66, and 63 for FH, EP, SMP, and MC, respectively, except that *N*'s were lower for Best scores because subjects sometimes failed to mark their Best response.

[b] No "Best" score was obtained for EP.

Table 7

Coefficient Alpha Reliabilities of Six-Item TST Scores[a]

Test	Study	Score					
		Best	Mean	Highest	Number	Unusual	Unusual-High
FH	Follow-up	.57	.54	.68	.57	-.08	.31
	Original	.48	.58	.60	.64	.47	.40
EP	Follow-up	--[b]	.53	.27	.62	.52	.36
	Original	--	.58	.48	.72	.49	.40
SMP	Follow-up	.30	.45	.61	.63	.54	-.06
	Original	.22	.41	.28	.59	.37	-.11
MC	Follow-up	.77	.76	.71	.81	.18	.47
	Original	.65	.74	.68	.77	.32	.10

[a] *N*'s are given in Table 6.

[b] No "Best" score was obtained for EP.

Table 8

Correlations of GRE and TST Scores

Test	Score	GRE Scores									
		V		Q		Adv.		Exp.		Soc.-Pers.	
		Follow-up	Original	Follow-up	Original	Follow-up	Original	Follow-up	Original	Follow-up	Original
FH	Best	.10	.34	.37	.34	.08	.28	-.02	.24	.16	.25
	Mean	.19	.31	.28	.28	.10	.24	-.02	.20	.19	.24
	High	.29	.38	.49	.37	.26	.31	.17	.26	.35	.31
	Num.	.37	.25	.57	.35	.32	.22	.35	.21	.32	.17
	Unus.	.11	.13	.05	.20	.12	.12	.10	.13	.19	.09
	U-H	.16	.19	.17	.19	.08	.17	.02	.14	.11	.16
EP[a]	Mean	.35	.37	.11	.31	.31	.33	.26	.29	.30	.31
	High	.42	.47	.27	.40	.41	.42	.36	.37	.36	.39
	Num.	.36	.35	.25	.27	.21	.29	.23	.27	.16	.22
	Unus.	.28	.13	.11	.09	.20	.07	.24	.09	.13	.01
	U-H	.20	.15	-.06	.13	.26	.11	.45	.15	.01	.03
SMP	Best	.34	.39	.48	.32	.36	.40	.34	.33	.37	.42
	Mean	.60	.39	.45	.34	.52	.45	.37	.36	.60	.46
	High	.65	.49	.48	.39	.57	.53	.51	.50	.53	.46
	Num.	.27	.15	.07	.14	.17	.13	.19	.19	.08	.02
	Unus.	-.24	.01	-.02	.03	-.17	-.09	-.07	-.03	-.29	-.14
	U-H	-.04	.20	.31	.21	.01	.18	.04	.21	-.10	.10
MC	Best	.03	.28	.26	.30	.23	.24	.19	.24	.27	.18
	Mean	.01	.29	.49	.34	.01	.25	.02	.24	.01	.18
	High	.03	.35	.45	.33	.07	.31	.04	.25	.14	.29
	Num.	.14	.22	-.11	.05	.19	.19	.14	.10	.29	.27
	Unus.	.35	.28	-.12	.10	.36	.23	.27	.18	.45	.26
	U-H	.27	.24	.23	.21	.17	.32	.13	.26	.22	.32

[a]No "Best" score was obtained for EP.

Table 9

Correlations of Questionnaire Variables with GRE Scores[a]

	Questionnaire Variable	V	Q	Adv.	Exp.	Soc.-Pers.
	Student/Dept. Quality					
2.	Plans PhD rather than MA	.19	.12	.18	.17	.17
3.	Undergraduate GPA	.22	.14	.20	.18	.19
4.	Graduate GPA	.18	.15	.26	.22	.23
5.	Attends accredited dept.	.38	.27	.34	.32	.29
29.	Rates dept. high in quality	.27	.23	.18	.18	.14
30.	Satisfied with department	.17	.14	.10	.09	.10
31.	Department quality index	.37	.31	.30	.28	.25
32.	Support through psychology	.22	.17	.27	.29	.19
33.	Support through fellowship	.27	.15	.17	.13	.19
	Program Emphasis					
6.	Plans academic program	.14	.15	.11	.20	-.01
27.	Program objective: practice	-.19	-.17	-.23	-.29	-.12
28.	Deemphasizes teaching	.03	.02	.04	.02	.04
	Interest Areas					
7.	Number of areas	.12	.01	.13	.12	.11
8.	Administration	-.06	.00	-.01	-.01	-.02
9.	Applied research	.19	.12	.16	.18	.10
10.	Basic research	.14	.12	.17	.23	.06
11.	Clinical practice	.01	-.06	.00	-.08	.13
12.	Guidance and counseling	-.15	-.21	-.22	-.23	-.14
13.	Teaching	.09	.06	.16	.16	.13
	Preferred Area					
14.	Administration	-.08	-.02	-.07	-.05	-.09
15.	Applied research	.06	.02	.02	.06	-.04
16.	Basic research	.10	.10	.14	.19	.07
17.	Clinical practice	-.03	-.03	-.03	-.11	.08
18.	Guidance and counseling	-.13	-.13	-.17	-.17	-.13
19.	Teaching	.03	.05	.05	.06	.02
	Self-Appraisal					
20.	Mean rating	.04	.03	.17	.18	.13
21.	Knowledge of psychology	-.02	-.08	.14	.17	.07
22.	Knowledge of statistics	.07	.21	.15	.16	.07
23.	Experimental design	.05	.01	.11	.13	.08
24.	Research interpretation	.15	.12	.23	.25	.15
25.	Clinical ability	-.10	-.17	-.12	-.20	.02
26.	Teaching ability	-.02	-.03	.08	.07	.07
	Professional Activities					
34.	Number of activities	.01	.02	.13	.19	.04
35.	Meetings, subscriptions	-.12	-.17	-.02	-.03	-.01
36.	Publications	.04	.00	.16	.18	.12

Table 9

Questionnaire Variable	V	Q	Adv.	Exp.	Soc.-Pers.
37. Planned, did indep. research	-.03	-.05	.01	.04	-.01
38. Did collaborative research	.06	.04	.13	.17	.05
39. Taught undergraduates	.07	.03	.11	.13	.07
40. Advised on statistics	.01	.09	.02	.04	-.03
41. Helped prepare book	.12	.05	.05	.05	.03
42. Worked with equipment	.00	.13	.09	.18	-.04
Other					
1. Has MA rather than BA	-.07	-.19	.11	.06	.14

[a]For correlations with V and Q scores, N,s are usually about 525, and for the Advanced Psychology Test scores they are usually about 650. With N of 525, a correlation of .09 is significant at the 5% level, and one of .11 is significant at the 1% level. Coefficients of .25 or higher are underlined.

Table 10

Correlations of Best Quality Scores with Questionnaire Variables[a]

	Questionnaire Variable	FH	EP[b]	SMP	MC
	Student/Dept. Quality				
2.	Plans PhD rather than MA	.14	--	-.08	-.07
3.	Undergraduate GPA	-.10	--	.27	-.04
4.	Graduate GPA	-.08	--	.03	.01
5.	Attends accredited dept.	.01	--	.03	.07
29.	Rates dept. high in quality	-.11	--	.15	-.05
30.	Satisfied with department	-.20	--	.02	-.06
31.	Department quality index	.03	--	.32	-.01
32.	Support through psychology	.15	--	.14	.07
33.	Support through fellowship	.08	--	-.08	-.06
	Program Emphasis				
6.	Plans academic program	-.09	--	.10	-.04
27.	Program objective: practice	-.02	--	-.09	.10
28.	Deemphasizes teaching	-.08	--	-.10	-.07
	Interest Areas				
7.	Number of areas	.01	--	.22	-.00
8.	Administration	-.15	--	.03	.03
9.	Applied research	-.08	--	.16	-.07
10.	Basic research	.02	--	.09	-.01
11.	Clinical practice	.25	--	.01	.04
12.	Guidance and counseling	-.13	--	-.23	.03
13.	Teaching	.01	--	.21	.15
	Preferred Area				
14.	Administration	.00	--	-.18	-.04
15.	Applied research	-.13	--	-.14	-.16
16.	Basic research	.10	--	.12	.17
17.	Clinical practice	.14	--	.03	.16
18.	Guidance and counseling	-.11	--	-.03	-.19
19.	Teaching	-.08	--	.14	-.05
	Self-Appraisal				
20.	Mean rating	.18	--	-.10	-.00
21.	Knowledge of psychology	-.11	--	-.11	-.09
22.	Knowledge of statistics	.26	--	-.28	.02
23.	Experimental design	.12	--	-.02	.07
24.	Research interpretation	.09	--	.14	.10
25.	Clinical ability	.35	--	-.23	.04
26.	Teaching ability	-.13	--	.07	-.17
	Professional Activities				
34.	Number of activities	.02	--	-.08	.09
35.	Meetings, subscriptions	-.01	--	-.06	-.01
36.	Publications	.15	--	-.10	-.03

Table 10 (continued)

	Questionnaire Variable	FH	EP[b]	SMP	MC
37.	Planned, did indep. research	.03	--	-.05	-.00
38.	Did collaborative research	-.01	--	-.09	.10
39.	Taught undergraduates	-.14	--	.12	.21
40.	Advised on statistics	.17	--	-.07	.02
41.	Helped prepare book	-.08	--	.03	.04
42.	Worked with equipment	.01	--	-.08	.03
	Other				
1.	Has MA rather than BA	-.00	--	-.04	-.18

[a]Correlations based on 3-item tests. Coefficients significant at the 5% level are underlined.

[b]No "Best" score was obtained for EP.

Table 11

Correlations of Mean Quality Scores
with Questionnaire Variables[a]

	Questionnaire Variable	FH	EP	SMP	MC
	Student/Dept. Quality				
2.	Plans PhD rather than MA	.08	.31	-.20	-.00
3.	Undergraduate GPA	-.04	-.10	.27	.07
4.	Graduate GPA	.03	.32	.03	-.02
5.	Attends accredited dept.	.06	.12	.09	.09
29.	Rates dept. high in quality	-.11	.01	.14	-.00
30.	Satisfied with department	-.22	.08	.17	.03
31.	Department quality index	.13	.04	.17	.11
32.	Support through psychology	.02	.04	.07	.06
33.	Support through fellowship	.10	.10	-.19	.10
	Program Emphasis				
6.	Plans academic program	-.04	-.06	-.03	.02
27.	Program objective: practice	-.12	.04	.01	.03
28.	Deemphasizes teaching	-.15	-.22	.04	.14
	Interest Areas				
7.	Number of areas	-.08	-.05	-.01	-.02
8.	Administration	-.10	-.03	.09	.05
9.	Applied research	-.16	-.25	.07	-.02
10.	Basic research	.03	.17	-.07	-.07
11.	Clinical practice	.02	.10	.01	.05
12.	Guidance and counseling	.12	.14	-.18	.05
13.	Teaching	-.08	-.05	-.05	-.04
	Preferred Area				
14.	Administration	-.02	-.28	-.02	.01
15.	Applied research	-.13	-.15	-.11	-.20
16.	Basic research	.16	.10	-.15	.11
17.	Clinical practice	.02	.07	.11	.16
18.	Guidance and counseling	.09	.19	-.04	-.08
19.	Teaching	-.14	-.01	.19	-.06
	Self-Appraisal				
20.	Mean rating	.10	-.01	.02	-.17
21.	Knowledge of psychology	-.07	-.13	-.00	-.10
22.	Knowledge of statistics	.31	.06	-.03	-.06
23.	Experimental design	-.09	-.06	-.01	-.12
24.	Research interpretation	.07	.05	.12	-.07
25.	Clinical ability	.18	.10	-.17	-.09
26.	Teaching ability	-.12	-.03	.11	-.20
	Professional Activities				
34.	Number of activities	-.06	-.05	-.06	-.01
35.	Meetings, subscriptions	-.09	.10	-.05	-.17
36.	Publications	.08	-.15	.00	.01

Table 11 (continued)

Questionnaire Variable	FH	EP	SMP	MC
37. Planned, did indep. research	-.02	-.05	-.06	-.00
38. Did collaborative research	-.10	-.14	-.14	.07
39. Taught undergraduates	-.15	-.00	.21	.06
40. Advised on statistics	.12	-.09	-.04	-.00
41. Helped prepare book	.06	.15	-.02	-.01
42. Worked with equipment	.00	-.02	-.15	.02
Other				
1. Has MA rather than BA	.08	.00	.01	-.17

[a]Correlations based on 3-item tests. Coefficients significant at the 5% level are underlined.

Table 12

Correlations of Highest Quality Scores with Questionnaire Variables[a]

	Questionnaire Variable	FH	EP	SMP	MC
	Student/Dept. Quality				
2.	Plans PhD rather than MA	.22	.33	-.07	.08
3.	Undergraduate GPA	-.12	.18	.12	.03
4.	Graduate GPA	-.02	.13	-.01	.07
5.	Attends accredited dept.	.01	.30	.25	.12
29.	Rates dept. high in quality	-.11	-.07	.13	-.04
30.	Satisfied with department	-.31	.04	.10	-.00
31.	Department quality index	.11	.05	.22	.10
32.	Support through psychology	.18	-.01	.16	.11
33.	Support through fellowship	.16	-.03	-.44	.24
	Program Emphasis				
6.	Plans academic program	-.08	.06	.09	-.11
27.	Program objective: practice	-.00	-.07	-.07	.05
28.	Deemphasizes teaching	-.15	.15	-.08	.14
	Interest Areas				
7.	Number of areas	-.06	-.01	.02	.10
8.	Administration	-.11	.01	.02	.03
9.	Applied research	-.20	.09	.03	.09
10.	Basic research	.02	.12	.07	-.15
11.	Clinical practice	.13	.06	-.02	.18
12.	Guidance and counseling	-.02	-.09	-.22	-.00
13.	Teaching	.05	-.02	.08	.05
	Preferred Area				
14.	Administration	.00	-.03	.02	-.04
15.	Applied research	-.17	-.23	-.06	-.14
16.	Basic research	.05	.24	-.07	-.09
17.	Clinical practice	.10	-.00	-.03	.25
18.	Guidance and counseling	-.03	-.06	.00	-.17
19.	Teaching	.04	.10	.17	.07
	Self-Appraisal				
20.	Mean rating	.02	.22	.11	-.14
21.	Knowledge of psychology	-.22	.16	.12	-.09
22.	Knowledge of statistics	.27	.12	-.02	-.05
23.	Experimental design	-.10	.19	.17	-.07
24.	Research interpretation	-.05	.12	.19	-.18
25.	Clinical ability	.22	.02	-.16	.03
26.	Teaching ability	-.03	.18	.02	-.14
	Professional Activities				
34.	Number of activities	-.04	.06	.08	.05
35.	Meetings, subscriptions	-.06	-.30	.09	-.04
36.	Publications	.00	.07	.11	.02

Table 12 (continued)

Questionnaire Variable	FH	EP	SMP	MC
37. Planned, did indep. research	-.07	.17	-.07	-.01
38. Did collaborative research	-.15	.22	-.04	.16
39. Taught undergraduates	.00	.04	<u>.25</u>	-.03
40. Advised on statistics	.20	.02	.07	.07
41. Helped prepare book	.05	-.03	.04	.07
42. Worked with equipment	-.03	.09	-.12	.04
Other				
1. Has MA rather than BA	.11	.20	.19	-.12

[a]Correlations based on 3-item tests. Coefficients significant at the 5% level are underlined.

Table 13

Correlations of Number of Responses Scores with Questionnaire Variables[a]

	Questionnaire Variable	FH	EP	SMP	MC
	Student/Dept. Quality				
2.	Plans PhD rather than MA	.19	.22	.17	.15
3.	Undergraduate GPA	-.01	.16	-.21	.03
4.	Graduate GPA	.05	-.07	-.02	.15
5.	Attends accredited dept.	-.09	.30	.17	.08
29.	Rates dept. high in quality	-.04	-.05	.00	.08
30.	Satisfied with department	-.18	-.07	.02	-.00
31.	Department quality index	.10	.24	.25	.18
32.	Support through psychology	.16	.16	.26	.08
33.	Support through fellowship	.05	.19	-.29	.17
	Program Emphasis				
6.	Plans academic program	-.08	.12	.17	-.18
27.	Program objective: practice	.12	-.16	-.19	-.06
28.	Deemphasizes teaching	.04	.27	-.30	.07
	Interest Areas				
7.	Number of areas	.10	.04	.03	.15
8.	Administration	.09	-.05	.06	-.01
9.	Applied research	.05	.00	-.01	.19
10.	Basic research	-.11	.15	.06	-.16
11.	Clinical practice	.17	-.07	-.11	.25
12.	Guidance and counseling	-.16	.11	.03	-.05
13.	Teaching	.26	.00	.04	.04
	Preferred Area				
14.	Administration	.04	.03	.09	-.05
15.	Applied research	-.05	-.21	.05	.19
16.	Basic research	-.12	.26	.05	-.33
17.	Clinical practice	.03	-.10	-.07	.15
18.	Guidance and counseling	-.06	.02	.10	-.17
19.	Teaching	.22	.13	-.15	.11
	Self-Appraisal				
20.	Mean rating	-.05	.12	.18	.03
21.	Knowledge of psychology	-.21	.23	.22	.04
22.	Knowledge of statistics	-.01	-.03	.03	-.03
23.	Experimental design	.11	.20	.24	.02
24.	Research interpretation	-.08	.03	.13	-.16
25.	Clinical ability	.01	-.07	.02	.17
26.	Teaching ability	.02	.04	-.03	.10
	Professional Activities				
34.	Number of activities	.06	.15	.32	.15
35.	Meetings, subscriptions	.00	-.05	.08	.23
36.	Publications	-.05	-.03	.23	.10

Table 13 (continued)

Questionnaire Variable	FH	EP	SMP	MC
37. Planned, did indep. research	-.02	.24	.21	-.06
38. Did collaborative research	.01	<u>.29</u>	.18	.24
39. Taught undergraduates	.15	-.02	.23	-.11
40. Advised on statistics	.11	.04	.21	.16
41. Helped prepare book	.03	-.17	.12	.10
42. Worked with equipment	.01	.19	.03	.05
Other				
1. Has MA rather than BA	-.13	-.03	.22	.09

[a]Correlations based on 3-item tests. Coefficients significant at the 5% level are underlined.

Table 14

Correlations of Number of Unusual Responses with Questionnaire Variables[a]

	Questionnaire Variable	FH	EP	SMP	MC
	Student/Dept. Quality				
2.	Plans PhD rather than MA	.22	.26	.21	.03
3.	Undergraduate GPA	.08	.09	-.25	-.06
4.	Graduate GPA	.01	-.24	-.05	.19
5.	Attends accredited dept.	-.09	.27	-.15	.04
29.	Rates dept. high in quality	-.09	.12	-.12	.06
30.	Satisfied with department	-.14	-.06	-.14	.02
31.	Department quality index	-.29	.25	-.03	.18
32.	Support through psychology	.12	.30	.14	.04
33.	Support through fellowship	-.08	.12	-.22	.01
	Program Emphasis				
6.	Plans academic program	.02	.23	.06	-.03
27.	Program objective: practice	-.12	-.11	-.07	-.02
28.	Deemphasizes teaching	.07	.20	-.23	.21
	Interest Areas				
7.	Number of areas	.00	.17	.00	.13
8.	Administration	.14	.17	-.04	-.03
9.	Applied research	-.01	.19	-.05	.04
10.	Basic research	-.03	.13	.06	-.02
11.	Clinical practice	-.15	-.03	-.15	.16
12.	Guidance and counseling	-.03	-.09	.11	-.11
13.	Teaching	.19	.00	.11	.13
	Preferred Area				
14.	Administration	.11	-.06	.02	.08
15.	Applied research	-.10	.03	.03	-.01
16.	Basic research	-.10	.25	.08	-.15
17.	Clinical practice	-.04	-.08	-.07	.03
18.	Guidance and counseling	.07	-.19	.05	-.26
19.	Teaching	.19	-.05	-.07	.35
	Self-Appraisal				
20.	Mean rating	.17	.14	.08	.14
21.	Knowledge of psychology	.16	.28	.06	-.02
22.	Knowledge of statistics	-.04	-.10	.07	.03
23.	Experimental design	.17	.22	.08	.20
24.	Research interpretation	.12	.09	.07	.07
25.	Clinical ability	.10	-.11	.06	.06
26.	Teaching ability	.02	.12	-.02	.23
	Professional Activities				
34.	Number of activities	.30	.19	.24	.23
35.	Meetings, subscriptions	.06	.16	.23	.09
36.	Publications	.23	.06	.26	.17

Table 14 (continued)

	Questionnaire Variable	FH	EP	SMP	MC
37.	Planned, did indep. research	.16	.19	-.06	.08
38.	Did collaborative research	.12	.21	.11	.28
39.	Taught undergraduates	.14	-.03	.14	-.03
40.	Advised on statistics	.03	.00	.10	.12
41.	Helped prepare book	.14	-.24	.01	.12
42.	Worked with equipment	.34	.17	.13	.15
	Other				
1.	Has MA rather than BA	-.26	.09	.07	.01

[a]Correlations based on 3-item tests. Coefficients significant at the 5% level are underlined.

Table 15

Correlations of Number of Unusual-High Quality Responses Scores with Questionnaire Variables[a]

	Questionnaire Variable	FH	EP	SMP	MC
	Student/Dept. Quality				
2.	Plans PhD rather than MA	.16	.12	.18	.17
3.	Undergraduate GPA	.19	.20	-.04	.16
4.	Graduate GPA	.21	.07	.05	.03
5.	Attends accredited dept.	-.11	.15	.03	.09
29.	Rates dept. high in quality	-.18	.11	.12	-.00
30.	Satisfied with department	-.24	.04	-.03	.02
31.	Department quality index	-.16	.16	.26	.33
32.	Support through psychology	.15	.29	.15	.01
33.	Support through fellowship	-.20	-.01	-.10	.23
	Program Emphasis				
6.	Plans academic program	.05	.14	.18	.19
27.	Program objective: practice	-.15	-.26	-.21	-.16
28.	Deemphasizes teaching	-.08	-.02	-.20	.24
	Interest Areas				
7.	Number of areas	-.07	.27	-.13	.10
8.	Administration	.03	.09	-.17	-.06
9.	Applied research	-.02	.21	-.06	.01
10.	Basic research	-.08	.32	.05	.03
11.	Clinical practice	-.14	.11	-.12	-.03
12.	Guidance and counseling	-.08	-.06	-.05	-.10
13.	Teaching	.17	-.01	-.02	.23
	Preferred Area				
14.	Administration	-.19	-.10	.02	-.11
15.	Applied research	-.01	.02	-.12	-.11
16.	Basic research	-.02	.18	.13	.16
17.	Clinical practice	.08	.01	-.07	-.06
18.	Guidance and counseling	.04	-.16	.03	-.20
19.	Teaching	-.02	-.11	.07	.26
	Self-Appraisal				
20.	Mean rating	.23	.12	-.11	-.05
21.	Knowledge of psychology	.23	.33	-.05	-.11
22.	Knowledge of statistics	.16	-.08	.00	-.07
23.	Experimental design	.08	.05	-.10	.14
24.	Research interpretation	.21	.04	.01	.07
25.	Clinical ability	.02	-.03	-.25	-.23
26.	Teaching ability	-.01	.08	-.03	.05
	Professional Activities				
34.	Number of activities	.20	.11	.18	.15
35.	Meetings, subscriptions	.01	.13	.02	-.04
36.	Publications	.30	.01	.15	.07

Table 15 (continued)

Questionnaire Variable	FH	EP	SMP	MC
37. Planned, did indep. research	.12	.04	-.07	-.02
38. Did collaborative research	.12	-.01	.24	.24
39. Taught undergraduates	.02	.07	.18	.06
40. Advised on statistics	-.04	.01	.00	-.03
41. Helped prepare book	.10	-.15	-.03	-.02
42. Worked with equipment	.24	.15	.17	.22
Other				
1. Has MA rather than BA	-.23	.04	-.04	-.11

[a]Correlations based on 3-item tests. Coefficients significant at the 5% level are underlined.

Table 16

Average Correlations and Significance Levels
Combining Data over Four Tests[a]

Questionnaire Variable	Score		
	Number	Unusual	Unusual-High
Student/Dept. Quality			
2. Plans PhD rather than MA	.18**	.18**	.16*
3. Undergraduate GPA			
4. Graduate GPA			
5. Attends accredited dept.			
29. Rates dept. high in quality			
30. Satisfied with department			
31. Department quality index	.20**		.17**
32. Support through psychology	.17**	.15*	.15*
33. Support through fellowship			
Program Emphasis			
6. Plans academic program			.14*
27. Program objective: practice			-.20**
28. Deemphasizes teaching			
Interest Areas			
7. Number of areas			
8. Administration			
9. Applied research			
10. Basic research			
11. Clinical practice			
12. Guidance and counseling			
13. Teaching			
Preferred Area			
14. Administration			
15. Applied research			
16. Basic research			
17. Clinical practice			
18. Guidance and counseling			
19. Teaching			
Self-Appraisal			
20. Mean rating		.13*	
21. Knowledge of psychology			
22. Knowledge of statistics			
23. Experimental design	.15*	.17**	
24. Research interpretation			
25. Clinical ability			-.13*
26. Teaching ability			
Professional Activities			
34. Number of activities	.18**	.24***	.16*
35. Meetings, subscriptions		.14*	
36. Publications		.18**	.13*

Table 16 (continued)

Questionnaire Variable	Score		
	Number	Unusual	Unusual-High
37. Planned, did indep. research			
38. Did collaborative research	.19**	.18**	.15*
39. Taught undergraduates			
40. Advised on statistics	.13*		
41. Helped prepare book			
42. Worked with equipment		.19**	.19**
Other			
1. Has MA rather than BA			

[a]Two-tailed probability levels.

*$p < .05$

**$p < .01$

***$p < .001$

APPENDIX

GRE QUESTIONNAIRE

1. Have you attended a graduate or professional school this academic year?

   1 ☐ Yes
   2 ☐ No - did not apply
   3 ☐ No - applied but was not accepted
   4 ☐ No - have been accepted and will begin attending in the future
   5 ☐ No - was accepted but did not attend
   6 ☐ No - attended previously but not this academic year
   7 ☐ No - other: _____

2. What kind of program have you attended or have you applied for?

   1 ☐ Psychology
   2 ☐ Education
   3 ☐ Other Social Science
   4 ☐ Professional School: _____
   5 ☐ Did not apply
   6 ☐ Other: _____

3. What is the highest degree you now hold?

   1 ☐ Bachelor's (B.A., B.S.)
   2 ☐ Master's (M.A., M.S., M.A.T.)
   3 ☐ Other: _____

4. What is the highest degree you plan to obtain?

   1 ☐ Bachelor's
   2 ☐ Master's
   3 ☐ Ph.D.
   4 ☐ Ed.D.
   5 ☐ D. Psy.
   6 ☐ Other: _____

5. Approximately what overall grade average did you receive for your undergraduate work?

   1 ☐ A
   2 ☐ A-
   3 ☐ B+
   4 ☐ B
   5 ☐ B-
   6 ☐ C+
   7 ☐ C
   8 ☐ C- or lower
   9 ☐ No grades

6. Approximately what overall grade average have you received for your graduate work to date?

1 ☐ A
2 ☐ A-
3 ☐ B+
4 ☐ B
5 ☐ B-
6 ☐ C+
7 ☐ C
8 ☐ C- or lower
9 ☐ No grades
10 ☐ Have not attended graduate school

THE REMAINING QUESTIONS ARE DIRECTED TO THOSE WHO HAVE ATTENDED GRADUATE PROGRAMS IN PSYCHOLOGY DURING THE CURRENT ACADEMIC YEAR. IF YOU ARE NOT IN THIS GROUP, PLEASE STOP HERE AND RETURN THE QUESTIONNAIRE IN THE ENCLOSED ENVELOPE. THANK YOU.

7. Which one of the following best describes your intended area of specialization within psychology?

1 ☐ Clinical
2 ☐ Cognitive
3 ☐ Counseling
4 ☐ Developmental
5 ☐ Educational
6 ☐ Experimental, Comparative, or Physiological
7 ☐ Measurement
8 ☐ Organizational, Personnel
9 ☐ Personality
10 ☐ School
11 ☐ Social
12 ☐ Other: _____

8. Assuming you could obtain exactly the position you would like, to which of the following activities would you devote significant amounts of your time? (Mark as many as apply.)

☐ a. Administration
☐ b. Applied Research
☐ c. Basic Research
☐ d. Clinical Practice
☐ e. Guidance and Counseling
☐ f. Teaching
☐ g. Other, in or related to psychology: _____
☐ h. Other, not related to psychology: _____

9. If forced to choose one, to which of these activities would you prefer to devote all your time?

1 ☐ Administration
2 ☐ Applied Research
3 ☐ Basic Research
4 ☐ Clinical Practice
5 ☐ Guidance and Counseling
6 ☐ Teaching

10. This question is concerned with skills and competencies within the field of psychology. We would like your assessment of your present level of achievement, relative to other students with a similar amount of graduate training in psychology. In each box place a number from 1 to 5, using the scale shown below.

1. Exceptionally well prepared
2. Above average
3. Average
4. Below average
5. No experience or does not apply

☐ a. General knowledge of psychological literature
☐ b. Knowledge of the literature in your area of specialization
☐ c. Familiarity with research techniques in your area
☐ d. Knowledge of psychometric and/or statistical theory
☐ e. Ability to apply statistics to research data
☐ f. Skill in conducting experiments involving human subjects
☐ g. Ability to gain insight into the problems of clients or patients
☐ h. Ability to design original research studies
☐ i. Ability to evaluate research designs
☐ j. Ability to teach complex ideas to undergraduates
☐ k. Ability to interpret research findings

11. Which of the following <u>most</u> characterizes the main objective of the program in which you are enrolled?

1 ☐ To prepare students to be competent researchers
2 ☐ To prepare students to be effective teachers
3 ☐ To prepare students to be effective practitioners in professional applications of psychological knowledge

12. Which of the following <u>least</u> characterizes the major objective of the program in which you are enrolled?

1 ☐ To prepare students to be competent researchers
2 ☐ To prepare students to be effective teachers
3 ☐ To prepare students to be effective practitioners in professional applications of psychological knowledge

13. Overall, how do you think your department is ranked on a national scale among graduate psychology departments?

1 ☐ Top 5%
2 ☐ Top 10%
3 ☐ Top 25%
4 ☐ Top 50%
5 ☐ Lower 50%

14. Is the university you are now attending:

1 ☐ Your top preference
2 ☐ One you are satisfied with, but not your top preference
3 ☐ One you are not really satisfied with

15. If you like, give the name of your university here:

_____

16. What has been your major source of financial support this year?

1 ☐ Fellowship
2 ☐ Assistantship or other employment in research
3 ☐ Assistantship or other employment in teaching
4 ☐ Other employment related to psychology
5 ☐ Other

17. Which of the following have you done within the current academic year? (Mark as many as apply.)

☐ a. Attended one or more meetings of a scholarly or professional society

☐ b. Subscribed to two or more scholarly or professional journals

☐ c. Been author or coauthor of a scientific paper accepted for presentation at a meeting of a scholarly or professional society

☐ d. Been author or coauthor of a scientific paper <u>submitted</u> for publication to a scholarly or professional journal

☐ e. Been author or coauthor of a scientific paper <u>accepted</u> for publication by a scholarly or professional journal

☐ f. Prepared a detailed proposal or plan for a dissertation, master's thesis, or other major research project

☐ g. Carried out an independent research project

☐ h. Carried out a research project in collaboration with another student or a faculty member

☐ i. Had teaching responsibility for a section of an undergraduate course

☐ j. Conducted a section of an undergraduate class on one or several occasions

(continued)

17. (continued)

- ☐ k. Frequently advised or tutored other graduate students on psychometric or statistical problems
- ☐ l. Assisted in editing of text or preparation of bibliographic material for a book
- ☐ m. Designed and built a piece of laboratory equipment
- ☐ n. Learned to operate or maintain a piece of electronic equipment
- ☐ o. Programmed a computer to analyze research data

18. We would appreciate any additional comments or explanations you have to make.

Thank you for your cooperation.